Robust adaptive predictive modeling and data deluge

Bogdan Gabrys
Smart Technology Research Centre,
Computational Intelligence Research Group,
Bournemouth University
bgabrys@bournemouth.ac.uk

Extended abstract of invited talk

We are currently experiencing an incredible, explosive growth in digital content and information. According to IDC [11], there currently exists over 2.7 zetabytes of data. It is estimated that the digital universe in 2020 will be 50 times as big as in 2010 and that from now until 2020 it will double every two years. Research in traditionally qualitative disciplines is fundamentally changing due to the availability of such vast amounts of data. In fact, data-intensive computing has been named as the fourth paradigm of scientific discovery [10] and is expected to be key in unifying the theoretical, experimental and simulation based approaches to science. The commercial world has also been transformed by a focus on BIG DATA with companies competing on analytics [12]. Data has become a commodity and in recent years has been referred to as the ‘new oil’.

There has been a lot of work done on the subject of intelligent data analysis, data mining and predictive modelling over the last 50 years with notable improvements which have been possible with both the advancements of the computing equipment as well as with the improvement of the algorithms [1]. However, even in the case of the static, non-changing over time data there are still many hard challenges to be solved which are related to the massive amounts, high dimensionality, sparseness or inhomogeneous nature of the data to name just a few.

What is also very challenging in today’s applications is the non-stationarity of the data which often change very quickly posing a set of new problems related to the need for robust adaptation and learning over time. In scenarios like these, many of the existing, often very powerful, methods are completely inadequate as they are simply not adaptive and require a lot of maintenance attention from highly skilled experts, in turn reducing their areas of applicability.

In order to address these challenging issues and following various inspirations coming from biology coupled with current engineering practices, we propose a major departure from the standard ways of building adaptive, intelligent predictive systems and moving somewhat away from the engineering maxim of “simple is beautiful” to biological statement of “complexity is not a problem” by utilising the biological metaphors of redundant but complementary pathways, interconnected cyclic processes, models that can be created as well as destroyed in easy way, batteries of sensors in form of pools of complementary approaches, hierarchical organisation of constantly optimised and adaptable components.

In order to achieve such high level of adaptability we have proposed a novel flexible architecture [5-6] which encapsulates many of the principles and strategies observed in adaptable biological systems. The main idea of the proposed architecture revolves around a certain degree of redundancy present at each level of processing represented by the pools of methods, multiple competitive paths (individual predictors), their flexible combinations and meta learning managing general population and ensuring both efficiency and accuracy of delivered solution while maintaining diversity for improved robustness of the overall system.

The results of extensive testing for many different benchmark problems and various snapshots of interesting results covering the last decade of our research will be shown throughout the presentation and a number of challenging real world problems including pollution/toxicity prediction studies [8-9], building adaptable soft sensors in process industry in collaboration with Evonik Industries [6-7] or forecasting demand for airline tickets covering the results of one of our collaborative research projects with Lufthansa Systems [3-4] will be discussed.

Given our experiences in many different areas we see that truly multidisciplinary teams and a new set of robust, adaptive tools are needed to tackle complex problems with intelligent data analysis, predictive modelling and visualisation already indispensible. It is also clear that complex adaptive systems and complexity science supported and driven by huge amounts of multimodal, multisource data will become a major endeavour in the 21st century.
References